4
Robustez do Reconhecimento de Voz

Nos capítulos anteriores foram apresentados a teoria e o funcionamento das etapas que conformam o sistema de reconhecimento de voz contínua. No entanto, se o que se deseja é um sistema que trabalhe de maneira robusta e confiável, como se tivesse sido treinado nas mesmas condições nas que se realiza o reconhecimento, é preciso acrescentar técnicas adicionais que não exijam uma carga computacional excessiva.

Na atualidade, estas técnicas de robustez constituem uma área de pesquisa fundamental no processamento de voz e estão divididas em três blocos principais, como foi mencionado no capítulo 1, as quais visam a gerar sistemas de reconhecimento de voz que apresentem elevadas taxas de acerto, mesmo em ambientes adversos.

Este capítulo apresenta o efeito do ruído sobre a sinal voz e as técnicas de robustez citadas anteriormente, e também descreve a estrutura e a distribuição dessas técnicas neste trabalho.

4.1
Reconhecimento de voz em presença de ruído

Uma das maiores dificuldades que existem nos sistemas reconhecedores de voz é a degradação do sinal quando a voz é contaminada em ambientes adversos. Isso se deve ao fato de que as condições ambientais nas que o sistema vai se desenvolver alteram as estatísticas dos vetores que as representam, obtendo-se desempenhos ruins devido as diferenças existentes entre as características de treinamento e as características de reconhecimento.

Uma das condições adversas mais relevantes que implica maiores dificuldades nos sistemas de reconhecimento de voz, é o ruído aditivo, tornando-se na atualidade o motor da investigação no campo de reconhecimento automático robusto de voz. Este tipo de ruído é adicionado ao sinal de voz no domínio do tempo, e pode ser considerado como:

– **Estacionário:** se possui uma densidade espectral que não varia com o tempo, como o caso do ruído aditivo branco, o qual tem um espetro de potência plana, ao contrário dos ruídos aditivos coloridos, onde o espetro de potência tem características diferentes para certas frequências.
– **Não estacionário:** se suas densidades espectrais mudam com o tempo,por exemplo, as vozes espontâneas, efeitos da respiração, etc.

Além desse tipo de ruído, existe um outro que da mesma maneira degrada o sinal de voz antes de ser registrado e processado. Esse tipo de ruído chamado de distorção, mistura-se de forma convolucional com o sinal de voz no domínio do tempo.

A Fig 4.1 mostra um modelo comumente utilizado para o ambiente acústico [45], onde é representado o problema geral de reconhecimento de voz em condições reais.

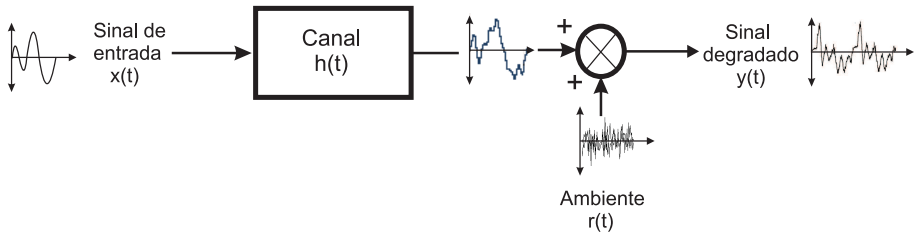

Figura 4.1: Diagrama de blocos do modelo de ambiente acústico.

onde h(t) representa o ruido convolutivo do canal, e r(t) representa o ruído aditivo do ambiente.

Segundo o modelo apresentado, o sinal de voz degradado, $y(t)$, pode ser expresso em termos do sinal de voz limpa $x(t)$, através da equação 4-1, que analiticamente determina como a adição de ruídos degrada o sinal de voz.

$$y(t) = x(t) * h(t) + r(t) \tag{4-1}$$

onde $x(t)$ e $r(t)$ são estatisticamente independentes e $h(t)$ é considerado como um deslocamento nos parâmetros da voz limpa sendo muitas vezes removido através de algum processo de filtragem linear [46].

Esta degradação do sinal diminui as taxas de reconhecimento nos ambientes reais, devido ao fato de que os testes recebem um sinal $y(n)$ bem diferente do que era esperado com $x(n)$. Esse descasamento entre as etapas de treinamento e teste provoca vários erros no reconhecimento, prejudicando o desempenho do sistema.

Segundo [47], uma solução a este problema é ter um banco de dados de treinamento igual ao número de condições adversas que possam ocorrer. Porém,

é uma tarefa difícil reunir dados de todos os ambientes possíveis. Devido a estas limitações, alguns trabalhos abordam o problema desenvolvendo técnicas robustas em cada um dos três blocos mencionados no capítulo 1, que são detalhados na seguinte seção.

4.2 Técnicas de robustez para o reconhecimento de voz em presença de ruído

Com o objetivo de gerar sistemas automáticos capazes de processar o sinal de voz emitido pelo ser humano e reconhecer a informação neste contida, com um alto rendimento de reconhecimento para ambientes adversos, foram desenvolvidos diversos métodos de robustez que reduzem o descasamento entre treino e teste.

Porém, é impossível fazer um listagem de todos os métodos propostos até hoje, pela grande quantidade de métodos encontrados na literatura. Por isso, como foi mencionado no capítulo 1, estes métodos foram agrupados em três técnicas importantes.

A seguir descrevem-se os aportes mais relevantes de cada uma destas técnicas.

4.2.1 Técnicas de realce de fala

As técnicas de realce de fala foram as primeiras a serem aplicadas nos métodos de robustez [48]. Visam eliminar o ruído do sinal antes de seu processamento e seu posterior reconhecimento, procurando obter a fala limpa a partir do sinal contaminado, como se mostra na Fig.4.2.

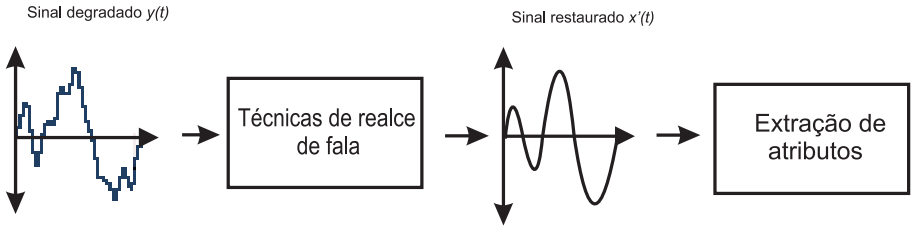

Figura 4.2: Restauração do sinal através de técnicas de de realce de fala

Estas técnicas são aplicadas no domínio espectral e baseiam-se no argumento de que a voz e o ruído são decorrelatados e são aditivos no domínio do tempo, pelo qual o espectro de potência do sinal degradado será a soma do espectro do sinal de voz e do ruído.

4.2.2
Técnicas de compensação de atributos

As técnicas de compensação de atributos atuam sobre as características parametrizadas, com o objetivo de recuperar o melhor possível os vetores de atributos limpos. Neste caso, os modelos acústicos treinados com voz limpa são utilizados para avaliar uma versão compensada da representação da voz, reduzindo os efeitos de descasamento entre as condições de referência e as de reconhecimento, como se mostra na Fig.4.3.

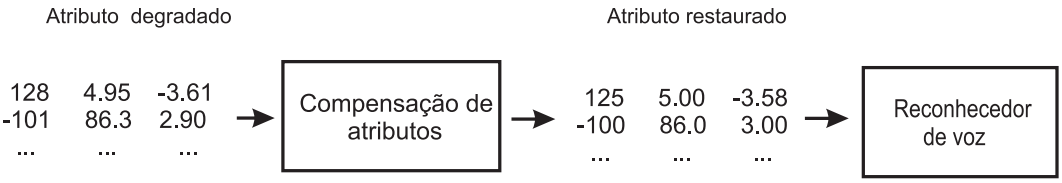

Figura 4.3: Restauração de atributos através de técnicas de compensação.

É importante considerar que dependendo do método aplicado para fazer a compensação de atributos, requer-se um treinamento específico para realizar o reconhecimento.

4.2.3
Técnicas de adaptação de modelos

Nestes tipos de técnicas tenta-se adaptar os reconhecedores às condições de ruído, isto é, adaptar os modelos acústicos obtidos no treinamento do sistema às condições de teste, visando avaliar o sinal de voz adquirida em condições de ruído com os modelos de voz ruidosos.

Uma característica importante destes tipos de técnicas é que se direciona o enfoque para os efeitos do ruído no sinal de voz, no lugar de tentar removê-lo. Isso faz com que os atributos do sinal de voz sejam resistentes ao ruído, sem necessidade de nenhum processamento adicional [49].

De acordo com as características que oferecem cada uma das técnicas apresentadas, escolheu-se avaliar as duas primeiras devido à sua eficiência computacional e capacidade de armazenamento menor, e são distribuídas como se apresenta na Fig. 4.4.

Nesta figura o bloco de pré-extração de atributos contém as técnicas de realce de fala e o bloco de pós-extração de atributos contém as técnicas de compensação de atributos. Cabe salientar que optou-se por implementar dois métodos de robustez para cada um dos blocos, a saber:

– Subtração Espectral e Wavelet Denoising para pre-extração de atributos.

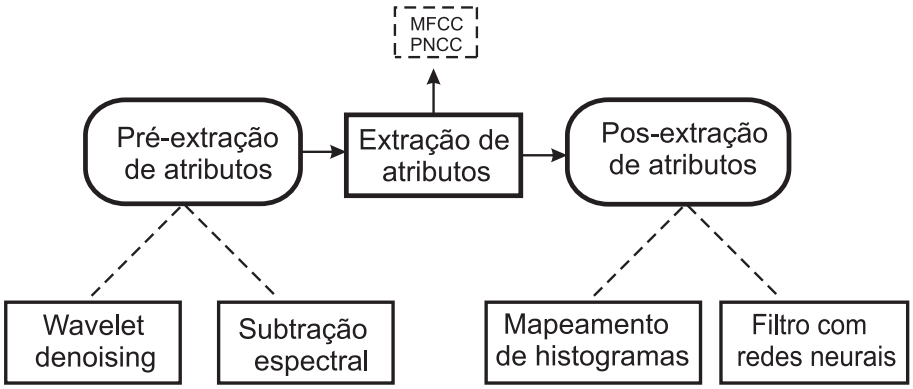

Figura 4.4: Distribuição das técnicas para o reconhecimento robusto de voz.

– Mapeamento de Histogramas e Filtro com redes neurais para pós-extração de atributos.

Os métodos serão agrupados e misturados, visando ter sistemas mais robustos e melhores taxas de reconhecimento. As implementações e resultados correspondentes a cada um destes métodos serão apresentados e analisados nos capítulos 5 e 6.